MP3 Re-evaluated

KEITH HOWARD LOOKS AT THE SOMETIMES SURPRISING RESULTS OF EVALUATING DIGITAL COMPRESSION CODECS

I wouldn't blame you if you thought that I was mad to be writing about MP3 at this juncture. It is, after all, a subject on which most audiophiles made their minds up years ago. My own view may encapsulate yours: that perceptual coding is an amazing technology (anything which can reduce the CD data rate by a factor of almost 12 to 128kb/s and still sound passably like the original is surely amazing) but, ultimately, lossy sounds as lossy does. In throwing away information, perceptual coding leaves fingerprints. And that's before we enter the realm of hi-res digital and the audible benefits that it brings.

So, why 'MP3 re-evaluated'? Well, it isn't me doing the re-evaluation, nor does the title of this piece mean that MP3 is either being lionised or subject to further vilification here. While it's my view that dynamic range compression has done more to damage music over the past two decades than file size compression, I'm no apologist for MP3, Dolby Digital, AAC, etc. No, it's the nature of the evaluation that is different here, and it has the potential for moving academic assessments of sound quality a step closer to the less formal (and more meaningful) ones that you and I make whenever we evaluate a recording or an item of hi-fi equipment.

Audio's objective/subjective divide, which has often been of chasmic proportions over the past 40 years, has always been characterised by a difference in language as well as a difference in approach. And yet, of course, science investigates subjective phenomena daily and does its best to forge causal links between objective fact and human perception.

You might find that difficult to credit, though, were you to read most of the assessments of perceptual codec performance published in the scientific literature. *Fig 1* (reproduced with permission from the European Broadcasting Union [EBU]) shows a typical outcome of such research[1], in this case a comparison of the multichannel performance of three different codecs – aacPlus (*aka* HE-AAC), Dolby AC-3 and Windows Media Audio (WMA) – operating at different data dates, performed by the German Institut für Rundfunktechnik (IRT) in 2004. The Mean Opinion Score (MOS) from subjective assessments using seven different signals is plotted on the vertical axis of the bar chart; the eighth grouping (far right) is the average combined score. The higher the score, the better the sound quality.

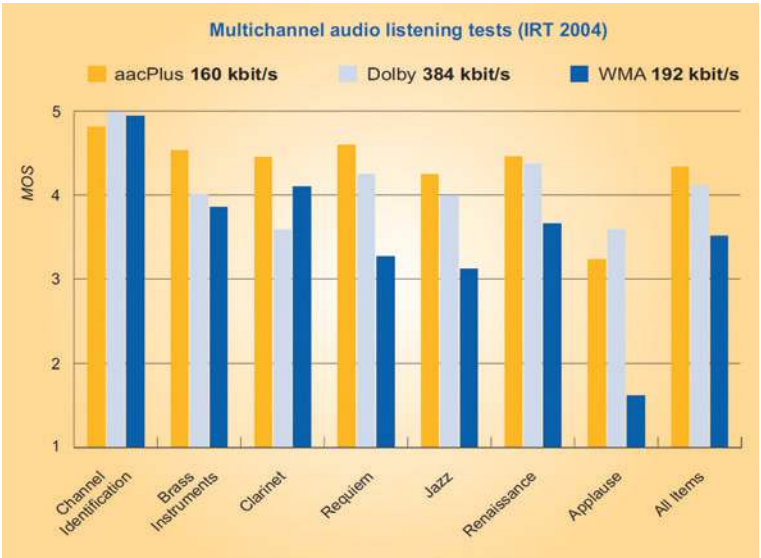

Figure 1. Results of a multichannel audio codec assessment conducted by IRT in 2004 (reproduced with permission)

As this shows, the performance of each codec changes according to the signal, as does its performance relative to the other two. This characteristic of perceptual codecs is such that researchers experienced in their subjective evaluation can often guess correctly a codec's identity from the degradations or artefacts it introduces on a particular type of signal. (Older audiophiles may smile at the relatively poor performance of all three codecs on applause, which for decades prior to the existence of perceptual coding was already known to be an unusually severe test of system sound quality.)

The five-point scoring scale used here, by the way, is defined as:

Grade	Impairment
5.0	Imperceptible
4.0	Perceptible but not annoying
3.0	Slightly annoying
2.0	Annoying
1.0	Very annoying

There are alternative scoring methods, particularly the continuous quality scale (CQS), which runs from 0 to 100 with 20-point spans characterised successively as Bad, Poor, Fair, Good and Excellent.